



Science Arts & Métiers (SAM)

is an open access repository that collects the work of Arts et Métiers Institute of Technology researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: <https://sam.ensam.eu>
Handle ID: <http://hdl.handle.net/10985/11352>

To cite this version :

Katy TCHA-TOKEY, Olivier CHRISTMANN, Emilie LOUP-ESCANDE, Simon RICHIR -
Proposition and Validation of a Questionnaire to Measure the User Experience in Immersive
Virtual Environments - International Journal of Virtual Reality - Vol. 16, n°1, p.33-48 - 2016

Proposition and Validation of a Questionnaire to Measure the User Experience in Immersive Virtual Environments

Katy Tcha-Tokey^{1*}, Olivier Christmann¹, Emilie Loup-Escande², Simon Richir¹

¹LAMPA, Arts et Métiers Paris Tech, France

²CRP-CPO, Université de Picardie Jules Verne, France

* Contact information:

Katy Tcha-Tokey
PHD Student
Arts et Métiers Paris Tech
4 rue de l'Ermitage
53000 Laval
Phone : (+33) 2 43 65 34 11
E-mail: Katy.tcha-tokey@ensam.eu

ABSTRACT

There are increasing new advances in Virtual Reality technologies as well as a rise in Immersive Virtual Environments research and in User eXperience research. Within this framework, we decided to address the overall user experience in Immersive virtual environments. Indeed, in our point of view, this topic is not fully dealt with in the scientific literature, neither in terms of user experience components nor in terms of user experience measurement methods. It is in this context that we conducted a study aiming at proposing and validating a unified questionnaire on User eXperience in Immersive Virtual Environment. Our questionnaire contains 10 scales measuring *presence*, *engagement*, *immersion*, *flow*, *usability*, *skill*, *emotion*, *experience consequence*, *judgement* and *technology adoption*. Scale construction was based on existing questionnaires. Our questionnaire was tested on 116 participants after they use the edutainment Virtual Environment “Think and Shoot”. The number of participants allows us to assess the reliability and the sensitivity of our questionnaire. Results show that 9 out of 10 subscales and 68 out of 87 items are reliable as demonstrated by an internal consistency analysis with Cronbach’s alpha and an item analysis. Findings also indicate that the scale scores from 6 subscales are considered normal distributed (e.g. presence) whereas the scale scores from 3 subscales are considered negatively skewed (e.g. skill).

KEYWORDS

Immersive Virtual Environment; User Experience; UX Model; Evaluation.

1. INTRODUCTION

After over 10 years of absence in the media landscape, Virtual Reality (VR) interest resumed in early 2012. Since, there is a rise in VR research to face the increase of new technology emergence. Nevertheless, according to us, the actual UX models for Virtual Environments (VE) discussed in scientific literature do not include the whole UX in VE key components.

This led us to propose a definition of the UX in Immersive Virtual Environments (IVE) that takes into account the multiple facets of the UX in several fields of VR (entertainment, education, edutainment).

Along with this new definition, we designed a new UX in IVE holistic model (*Figure 1*) and designed a suitable measurement method based on our model [23]. The model is based on the key components recommended by the literature and the designed measurement method is a questionnaire based on the components of our model. Our questionnaire is designed on the basis of existing questionnaires, since most of the UX in IVE components of our model can be measured through general UX questionnaires (non-VR specific) or specific UX for VR questionnaires.

This paper describes the questionnaire validation step. In fact, we want to make sure that the items selected to design our questionnaire measure properly the UX components from the original questionnaires. Firstly, we review the selected questionnaires we based our own questionnaire on. Secondly, we describe the experiment that took place to validate our questionnaire and finally we discuss the reliability and the sensitivity of our questionnaire.

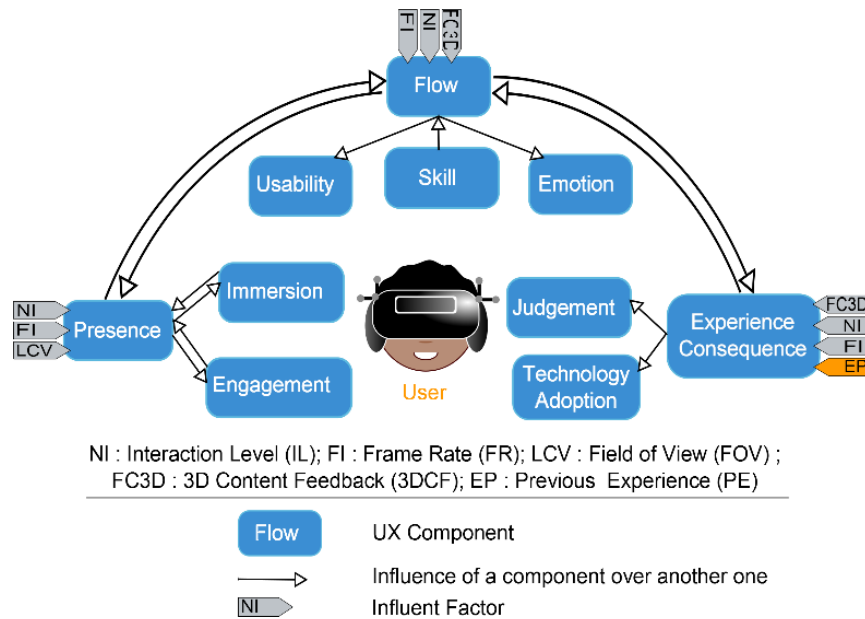


Figure 1. Our holistic User Experience in Immersive Virtual Environment model.

2. PROPOSITION OF A MODEL

The UX is defined by a variety of different components depending on the field the experience is lived in. In our study, we define the UX through 10 components. These components also structure our UX questionnaire in 10 subscales (APPENDIX 1). Our questionnaire was designed on the basis of 9 UX questionnaires. We define the 10 UX key components that compose our questionnaire in section 2.1. Two steps led us to our unified questionnaire: the UX questionnaires selection and the items selection. The questionnaires selection criteria are detailed in section 2.1. The items selection criteria and the internal structure of our questionnaire are detailed in section 2.2.

2.1 USER EXPERIENCE QUESTIONNAIRES REVIEW

The UX in IVE can be measured by either subjective methods or objective methods. Yet a combination of both methods might provide results that are more reliable [27]. Subjective methods (e.g. Questionnaire, interviews, focus groups...) provide results through the user's point of view, attitudes or preferences, whereas objective methods (e.g. Electroencephalogram, Electromyogram, time completion, level reached...) provide results through observable evidence. Questionnaire is, currently, the most commonly used method for the measure of UX components (e.g. presence, engagement, immersion, flow, emotion, judgment...), furthermore, a large number of questionnaires have been proven valid and reliable. Therefore, we base our approach on this fact and, thus, chose to focus on the questionnaire method in this paper. Actually, to the best of our knowledge, no UX questionnaire integrates all of the UX key components for IVE concern. In the second place, objective methods used today to assess user experience are questionable due to signal contamination : motor interference in Brain Computer Interaction (BCI) and Electromyogram (EMG) [17], signal distinction of two close emotions such as stress or excitement in Electrocardiogram (ECG) or Electroencephalogram (EEG) [4], [14], late response latency in Galvanic Skin Response (GSR) and Skin Temperature (SKT) 22[14]. A UX assessment tool based only on objective methods communicates meaningful data and a direct interpretation of user behavioral and physiological state [9]. Nevertheless, for a complete diagnosis of UX with the user's thoughts patterns and believes, it is preferable to combine the objective data with subjective data using suitable tools such as a questionnaire and "UX heat maps" [9]. It should be noted that this paper's objective is to present and validate a subjective method: our unified UX questionnaire. Once validated, our questionnaire's data might be combined at will with other objective data.

Various components are relevant to measure the UX. In a previous study [23], we propose 10 key components (*presence, engagement, immersion, flow, usability, skill, emotion, experience consequence, judgement and technology adoption*) all of which shapes the overall UX in IVE. The items assessing each component of the UX can be found in several existing questionnaires.

In order to help us choose the suitable questionnaires we defined three questionnaires selection criteria:

- Validity of the questionnaire (i.e. the whole validation process of the questionnaire is published in a paper).
- Frequent use of the questionnaire or if the questionnaire is based or inspired by a frequent used questionnaire (i.e. we consider frequent the use of a questionnaire which is cited at least by 20 other papers in the scientific literature).

After defining the criteria, we found a suitable questionnaire for each the UX components we propose. The suitable questionnaires and the UX components are defined below:

Presence is a component defined as the user's "sense of being there" in the VE. The concept of presence can be divided into two categories: physical presence in the virtual environment and social presence in the collective or collaborative virtual environment [20]. Most measures of presence try to address both. *Engagement* is a component defined as the "energy in action, the connection between a person and its activity consisting of a behavioral, emotional and cognitive form". The Presence Questionnaire (PQ) created by Witmer and Singer measures presence and engagement [26], it identifies the degree to which individuals experience presence and engagement in VE. This questionnaire is composed of 24 items divided in 5 subscales: involved/control, natural, auditory, resolution and interface quality. Items 4, 6, 10, 13, 20 actually measure the engagement component.

Immersion is a component defined as the “illusion” that “the virtual environment technology replaces the user’s sensory stimuli by the virtual sensory stimuli”. The Immersion Tendency Questionnaire (ITQ) created by Witmer and Singer measures immersion [26], it identifies the tendency of individuals to be immersed. This questionnaire is composed of 16 items divided in 3 subscales: involvement, focus and game.

Flow is a component defined as “a pleasant psychological state of sense of control, fun and joy” that the user feels when interacting with the VE. The Flow4D16 questionnaire created by Heutte measures the flow component [12]. It identifies the degree with which the user is absorbed by his task. The questionnaire consists of 16 items divided in 4 subscales: cognitive absorption, altered time perception, lack of self-preoccupation, well-being.

Skill is a component defined as the knowledge the user gain in mastering his activity in the virtual environment. The Computer Self-Efficacy (CSE) questionnaire created by Murphy measures the skill component [18]. It identifies the attitude of a user toward a computer technology, the degree with which he feels comfortable with a computer. This questionnaire is a reference in the education field to evaluate adult student’s computer skills. The questionnaire consists of 32 items with 3 subscales representing different levels assessment of computer skills: beginning, advanced, mainframe.

Emotion is a component defined as the feelings (of joy, pleasure, satisfaction, frustration, disappointment, anxiety ...) of the user in the VE. The Achievement Emotions Questionnaire (AEQ) created by Pekrun measures the emotion component [21]. It identifies the emotion experienced in achievement situations. There are 3 subscales representing 3 situations: class-related, learning-related and test-related. This questionnaire is based on 9 emotions: enjoyment, hope, pride, relief, anger, anxiety, shame, hopelessness, boredom. The questionnaire consists of 232 items. It proposes a large number of situations that matches or that can be easily translated in a situation such as a user being in a VE.

Usability is a component defined as the ease of learning (learnability and memorizing) and the ease of using (efficiency, effectiveness and satisfaction) the VE. The System Usability Scale (SUS) created by Brooke measures the usability component [3]. This scale has been created on a base of 50 usability questionnaires. It identifies “the appropriateness of a purpose”, in other words, it identifies if the way we propose to use our VE is appropriate. The questionnaire consists of 10 items and is unidimensional.

Technology adoption is a component defined as the actions and decisions taken by the user for a future use or intention to use of the VE. The Unified Technology Acceptance and Use of Technology (UTAUT) questionnaire created by Venkatesh and al. measures the technology adoption component [25]. It identifies the degree with which the user will adopt and use the system, in other words, the likelihood of success for new technology introduction. This questionnaire consists of 31 items divided in 8 subscales: performance expectancy, effort expectancy, social influence, facilitating conditions, attitude toward using technology, self-efficacy, anxiety, behavioral intention to use the system.

Judgement is a component defined as the overall judgement of the experience in the VE. The AttracDiff questionnaire created by Hassenzahl, Burmester and Koller measures the judgement component [11]. It identifies the user’s attraction in a pragmatic and hedonic way towards the system. This questionnaire consists of 28 items divided in 4 subscales: perceived pragmatic quality, perceived hedonic quality-stimulation, perceived hedonic quality-identification, attractiveness.

Experience Consequence is a component defined as the symptoms (e.g. the “simulator sickness”, stress, dizziness, headache ...) the user can experience in the VE. The Simulator Sickness Questionnaire (SSQ) created by Kennedy measures the experience consequence component [13]. It identifies the negative consequences the user can have while using the IVE. These negative consequences are assessed through 16 items divided in 3 subscales: nausea, oculomotor problems, disorientation.

2.2 OUR UNIFIED UX QUESTIONNAIRE

2.2.1 STRUCTURING OF THE INSTRUMENT

The questionnaire we designed focuses on measuring the UX in IVE. It is composed of a set of items that gathers the user's opinion, beliefs and preferences on the VE that he/she experienced in terms of *presence*, *engagement*, *immersion*, *flow*, *usability*, *skill*, *emotion*, *experience consequence*, *judgement* and *technology adoption*. Our questionnaire is a unified questionnaire based on nine other existing questionnaires (PQ, ITQ, Flow4D16, CSE, AEQ, SUS, UTAUT, AttracDiff, SSQ). We suggest the use of our unified questionnaire to measure the UX in order to take into account the various facets of the UX in IVE. The idea is to offer to the user, after a certain amount of time in the VE, one unique questionnaire that measures the whole 10 components of the UX. As such questionnaire does not exist, we needed to elaborate our own [24] by choosing 3 items by subscale from each of the existing questionnaires. This questionnaire (*Table 1*) is composed of 10 subscales, 87 items among which 75 items have a 10-point Likert scale, 12 items (grouped in four) have a differential scale items and 3 items are open questions. The 87 items can be seen in APPENDIX 1.

Items
<i>Presence</i>
1. The virtual environment was responsive to actions that I initiated.
<i>Engagement</i>
6. The sense of moving around inside the virtual environment was compelling.
<i>Immersion</i>
16. I felt stimulated by the virtual environment.
<i>Flow</i>
23. I felt I could perfectly control my actions.
<i>Usability</i>
34. I thought the interaction devices (oculus headset, gamepad and/or keyboard) were easy to use.
<i>Emotion</i>
37. I enjoyed being in this virtual environment.
<i>Skill</i>
52. I felt confident selecting objects in the virtual environment.
<i>Judgement</i>
58. A. Personally, I would say the virtual environment is impractical/practical
<i>Experience consequence</i>
62. I suffered from fatigue during my interaction with the virtual environment.
<i>Technology adoption</i>
71. If I use again the same virtual environment, my interaction with the environment would be clear and understandable for me.

Table 1. English translation of some items used in our unified UX questionnaire.

We relied on a single criterion to select the 87 items of our questionnaire: the meaning of the chosen items had to be different enough from each other (even if they measure the same component), so that the user does not find the items redundant.

By choosing to keep at most 3 items by subscale, in the end, our questionnaire consists of 87 items. English translation of some items are presented in *Table 1*. This questionnaire comprised ten subscales as described below. Presence was assessed using 12 items (e.g. *“The virtual environment was responsive to actions that I initiated”*) adapted from PQ scales [26]. Engagement was assessed using 3 items (e.g. *“The sense of moving around inside the virtual environment was compelling”*) adapted from PQ scales [26]. Immersion was assessed using 7 items (e.g. *“I felt stimulated by the virtual environment”*) adapted from ITQ scales [26]. Flow was assessed using 11 items (e.g. *“I felt I could perfectly control my actions”*) adapted from Flow4D16 scales [12]. Usability was assessed using 3 items (e.g. *“I thought the interaction devices (oculus headset, gamepad and/or keyboard) was easy to use”*) adapted from SUS scales [3]. Emotion was assessed using 14 items (e.g. *“I enjoyed being in this virtual environment”*) adapted from AEQ scales [21]. Skill was assessed using 6 items (e.g. *“I felt confident selecting objects in the virtual environment”*) adapted from CSE scales [18]. Judgement was assessed using 12 items (e.g. *“Personally, I would say the virtual environment is impractical/practical”*) adapted from AttracDiff scales [11]. Experience consequence was assessed using 9 items (e.g. *“I suffered from fatigue during my interaction with the virtual environment”*) adapted from SSQ scales [13]. Technology adoption was assessed using 9 items (e.g. *“If I use again the same virtual environment, my interaction with the environment would be clear and understandable for me”*) adapted from UTAUT Scales [25]. We added 3 open questions at the end of the questionnaire to allow the user express the positive as well as the negative experience he wish to share and the improvements he wish to provide to the environment.

We made some arrangement in order to create a questionnaire better related to VE:

- In the PQ, some subscales (e.g. IFQUAL: Interface Quality; NATRL: Natural, AUD: Auditory, RESOL: Resolution) only had 2 or 3 items, in that case we did not have to make a selection and picked all of the items of the subscale (e.g. AUD: “14 - I correctly identified sounds produced by the virtual environment.”; “15 - I correctly localized sounds produced by the virtual environment.”).
- In the ITQ, the items in the subscale GAMES could hardly apply to VE and the items from the involvement (INVOL) and FOCUS subscales did not apply right away to our context. Therefore, we chose items that could easily be adjusted to our context, and rewrote the items if necessary (e.g. “How mentally alert do you feel at the present time?” becomes “16 - I felt mentally alert in the virtual environment.”, “How frequently do you find yourself closely identifying with the characters in a story line?” becomes “18 - I identified to the character I played in the virtual environment.” ...).
- In the AEQ, one subscale could hardly apply to VE (i.e. Relief). For the 2 remaining subscales we chose to select 3 items, one for each emotion category (positive activating: enjoyment, negative activating: anxiety, negative deactivating: boredom).
- In the UTAUT questionnaire, 5 direct determinants of intention subscales could hardly apply to VE or were redundant with other items already selected (i.e. performance expectancy, social influence, self-efficacy, anxiety, behavioral intention to use the system). 3 of the subscales did apply to VE (i.e. effort expectancy, attitude toward using technology, facilitating conditions).
- We adjust most of the items we selected so that they could fit perfectly to VE. In some cases, changing the words “system” or “class” to “virtual environment” was enough (e.g. “I enjoy being in class” becomes “37: I enjoyed being in the virtual environment”), in other cases we did adjust the whole item to apply for VE (e.g. “I feel confident making selections from an onscreen menu.” becomes “52 - I felt confident selecting objects in the virtual environment.”)

2.2.2 ANSWER MODALITIES AND SCORING

The participants UX scores were collected through a 10-point Likert scale (1 = strongly disagree, 10 = strongly agree) for 75 items. For the 12 items (grouped in 4) of the judgement scale scores were collected through a semantic differential scale: point 1 was coded as a negative-connoted adjective (e.g. impractical, confusing, amateurish ...) whereas point 10 was coded as a positive-connoted adjective (e.g. practical, clear, professional ...). A high score in a subscale means that the UX component measured is highly perceived by the participant (e.g. Presence equals to 9 means that the participant felt really present, “he felt there”, while he was in the Virtual Environment). A low score in a subscale means that the UX component measured is poorly perceived by the participant (e.g. Presence equals to 2 means that the participant did not really felt present, while he was in the Virtual Environment, there was few or no moments where he would easily forget about the real environment).

2.2.3 EXECUTION TIME

This questionnaire execution time varies from 15 to 20 minutes according to our observation of the participants.

2.2.4 USE CONDITIONS

Our UX questionnaire for IVEs is made available to help designers assess UX in IVE. Individual and research practices are welcome to use and adapt, if required, our UX questionnaire for IVEs for their own work, provided that acknowledgement is given and provided that it is not used for commercial services. Copies of the French version of our UX questionnaire for IVEs are available from the authors via e-mail.

3. VALIDATION OF OUR MODEL

We conducted experiments with the edutainment IVE prototype “Think and Shoot” and used our questionnaire to measure the UX for two categories of participants: the experienced and the non-experienced individuals in 3D technologies. We selected the commonly used psychometric properties to validate questionnaires: the reliability and the sensitivity.

3.1 AIM OF OUR STUDY

Our study aims at making a UX questionnaire available for VE designers and help them assess and improve the UX in VE prototypes or VE final products. Our goal, in this paper, is to validate our UX questionnaire for IVEs, designed using components and items of existing questionnaires detailed in section 2.1.

3.2 METHOD

3.2.1 PARTICIPANTS

116 participants (25 women and 91 men) aged 18-63 years ($M = 24.6$, $SD = 7.55$) took part in the study. 88 participants work or study in Information and Communications Technology (ICT) or Computer Science fields (e.g. VR engineers, VR research engineers, network administrator, web developer, web designer, graphic designer, PhD student in VR, master degree undergraduates in VR, professional degree undergraduates in ICT, technical degree undergraduates in Multimedia and Internet, ...). 19 participants work or study in various other fields (specialized education, marketing, food service, military, public relations, Bank ...). 9 participants did not answer the activity question.

79 participants are considered experienced with 3D technologies, they scored at least 11/42 points (see section 3.2.4 for scoring details) in the 3D technology expertise survey ($M = 19.99$, $SD = 7.29$), they use 3D technologies in a regular basis to play 3D video games or to create 3D contents. 37 participants are considered non-experienced, they scored less than 11/42 points in the 3D technology expertise survey ($M = 6.46$, $SD = 2.80$), they never use or use few 3D technologies dedicated to 3D contents or they never (or very little) play with 3D video games (Table 2). “11 /42” points corresponds to a hypothetical score of a participant that would have checked point 1 coded as “Little” for every technology we listed. He has some experience with all of the 3D technologies we listed in the survey.

Profile	Level of experience	Participants
Experienced	> or = 5 Points in the 3D technology expertise survey	79
Non-experienced	< 5 points in the 3D technology expertise survey	37

Table 2. Number of experienced versus non-experienced participants in 3D Technologies.

3.2.2 EXPERIMENTAL GROUPS

Before presenting the experimental groups, it should be noted that, although we present all the experimental groups and conditions in this study, for the need of this UX questionnaire validation, we will only be interested in the UX measures under the control_condition. First, because all the participants have tested this condition at least one time. Secondly, because we would like to validate the UX questionnaire in the same condition for all participants, to create more reliable results. Therefore, we will deal with the UX comparison under different conditions, in future studies.

Experimental groups (Table 3. *Experimental groups*Table 3) were created to examine the influence of external factors (e.g. field of view) on the UX. Actually, for the groups 1 to 4, the goal is to compare the UX measured under the control condition (i.e. control_condition) with the UX measured under the modified condition (e.g. modified_condition1). The other goal is to compare the UX measured on the experienced participants (group 5) with the UX measured on the non-experienced participants (group 6).

The participants were placed into 6 experimental groups.

Group 1 composed of 11 participants experienced in 3D technologies tested the control_condition and the modified_condition1 (i.e. same condition as the control_condition except that the field of view is at 32° [10]).

Group 2 composed of 11 participants (different from the previous group) experienced in 3D technologies tested the control_condition and the modified_condition2 (i.e. same condition as the control_condition except that the framerate is at 30 FPS). Group 3 composed of 11 participants (different from the previous groups) experienced in 3D technologies tested the control_condition and the modified_condition3 (i.e. same condition as the control_condition except that the interactivity device was a keyboard). Group 4 composed of 11 participants (different from the previous groups) experienced in 3D technologies tested the control_condition and the modified_condition4 (i.e. same condition as the control_condition except that there was no special feedback). Group 5 composed of 19 participants (different from the previous groups) experienced in 3D technologies tested the control_condition only. Group 6 composed of 53 participants (different from the previous groups) not experienced in 3D technologies tested the control_condition only.

Profile	Groups	Conditions	Participants
Experienced	1	control_condition = field of view: 106°; framerate: 70 FPS; interactivity device: gamepad; special feedback: minimap	11
		modified_condition1 = <i>field of view: 32°</i> ; framerate: 70 FPS; interactivity device: gamepad; special feedback: minimap	
Experienced	2	control_condition = field of view: 106°; framerate: 70 FPS; interactivity device: gamepad; special feedback: minimap	11
		modified_condition2 = field of view: 106°; <i>framerate: 30 FPS</i> ; interactivity device: gamepad; special feedback: minimap	
Experienced	3	control_condition = field of view: 106°; framerate: 70 FPS; interactivity device: gamepad; special feedback: minimap	11
		modified_condition3 = field of view: 106°; framerate: 70 FPS; <i>interactivity device: keyboard</i> ; special feedback: minimap	
Experienced	4	control_condition = field of view: 106°; framerate: 70 FPS; interactivity device: gamepad; special feedback: minimap	11
		modified_condition4 = field of view: 106°; framerate: 70 FPS; interactivity device: gamepad; <i>special feedback: no minimap</i>	
Experienced	5	control_condition = field of view: 106°; framerate: 70 FPS; interactivity device: gamepad; special feedback: minimap	19
Non-experienced	6		53

Table 3. Experimental groups of our study.

3.2.3 PROCEDURE

The experiment took place in the Presence & innovation team building called the “Ingénierium”. A 16 m² room (*Figure 2*) was rearranged and used only for the experiment purpose during two months (from the 1st of February of 2016 to the 31st of march 2016).



Figure 2. The 16m² room dedicated to the experiment purpose.

The experiment had three steps:

During the first step of the experiment, we installed the participants in the experiment room and asked them to read and sign a consent document presenting the laboratory and the experiment confidentiality rules. We then asked them to complete a “participant identification survey”.

During the second step of the experiment, first, we explained the whole experiment goal to the participants. Secondly, we explained the training goal and we asked the participants to put on the Oculus and the audio headsets for a training session of about 5 minutes (the participants could ask for more or less training time if they felt more or less comfortable in the IVE).

During the third step of the experiment, we explained the regular session goal to the participants. We then asked them to put on the Oculus and audio headsets for the regular session of 5 minutes. After the end of the session, the participants completed our UX questionnaire. Each participant spent between 30 to 45 minutes in the experiment room.

3.2.4 MATERIAL AND MEASURES

A consent document was used to inform the participant about the laboratory activity and to collect his agreement to participate in our experiment under the announced conditions (e.g. recorded experiment, confidentiality...). This document required the participant’s personal information (i.e. name, date of birth, address, occupation).

A participant identification survey was used to collect the user's skills. The participant's last diploma, and current diploma or occupation were asked. 3 items with a 5-point Likert scale were dedicated to programming expertise (0 = No knowledge, 4 = Excellent knowledge). Two multiple-choice questions were used to assess the participant's ability to recognize a function and a parameter in an instruction: one point was given for a good answer and zero point was given for a false answer.

Two matrix scale questions were dedicated to technology expertise (0 = Never, 1 = Little, 2 = Sometimes, 3 = Often, 4 = Always). The first matrix scale question was dedicated to the usage frequency of interaction devices such as VR headset, gamepad, joystick, Kinect, leap motion... and the second matrix scale question was dedicated to usage frequency of 3D video games and 3D software such as 3D scene design software (i.e. Virtools, Unity...), modelling software (i.e. 3DSmax, Maya ...), CAO software (i.e. AutoCAD, Architectural desktop ...).

Our UX in IVE questionnaire of 87 items and 3 open questions is used to assess the UX. All items and questions were originally in French. Our UX questionnaire (APPENDIX 1) consists of 12 items to measure presence, 3 items to measure engagement, 7 items to measure immersion, 11 items to measure flow, 3 items to measure usability, 6 items to measure skill, 15 items to measure emotion, 9 items to measure experience consequence, 12 items (grouped in 4) to measure judgement and 9 items to measure technology adoption.

The experiment consists of the edutainment VE prototype "Think and Shoot" designed with the development tool UNITY ©. The goal in the edutainment VE is to collect three types of balls and to shoot on two types of evil creatures pursuing the participants, according to the instructions given on a panel displayed inside the VE.



Figure 3. Virtual edutainment environment screen shot of a fire evil creature in the first level.

We proposed a training session and a regular session to the participants. In the training session, after collecting the balls, the participant earns a point if he shoots correctly a fire ball on the blue sphere target, an ice ball on the red sphere target and a lightning ball on the green sphere target.

In the regular session, after collecting the balls, the participant earns a point if he shoots correctly an ice ball on the fire evil creature, a fire ball on the ice evil creature. If he shoots a lightning ball on both evil creatures, they are then frozen and they cannot move forward anymore. There are six levels in the edutainment VE.

In the first level, the participant can only shoot ice balls and only one fire evil creature pursues him. In the second level, there are more fire evil creatures. In the third level, the participant can shoot both ice balls and lightning balls. In the fourth and fifth levels, the participant can shoot the three types of balls (ice, fire, and lightning) and he can shoot on both ice and fire evil creatures (*Figure 3*).

The edutainment VE “Think and Shoot” is displayed in an Oculus development kit 2 (DK2). To interact with the VE, a Logitech wireless gamepad is used. The 3D spatialized sound is rendered in a Tritton AX 180 audio headset. The VE is launched on a Dell 64bits with 4GB of RAM computer and an Intel® Xeon® processor, CPU E5-16030 2.80GHz. The computer operating system is Windows 10 Professional. Oculus runtime SDK 0.7 and NVIDIA 356.04 GeForce win10 drivers were installed.

The VE factors were fixed to a field of view of 106° and a framerate of 70 FPS (recommended by the Oculus Best Practices [28]). The Gamepad allowed the user to move forward, backward and on the sides with one of the joysticks, and with the other, he could rotate on himself. The user had a minimap of the environment to help him localize the balls.

3.3 DATA ANALYSIS

3.3.1 MAIN PSYCHOMETRIC PROPERTIES

Questionnaires are self-appraisal methods and thus induce two kinds of problems: the misunderstanding of the items’ meaning and the risk of giving a stereotypical answer. This is why we conducted this study to analyze the items and scales quality of our questionnaire through three recommended psychometric properties [8]: the reliability, the validity and the sensitivity.

The reliability is a psychometric quality that assesses the consistency of a measure. A highly reliable measure is a measure that produces the same result under consistent conditions. There are various types of reliability, for example the test-retest assessed by observing stability of the results throughout time or the internal consistency assessed by measuring the Cronbach’s alpha or the item correlation (through an item analysis).

The validity is a psychometric quality that assesses the accuracy of the affirmations that can be done through test scores. There are various types of validity, for example the construct validity assessed by analyzing the factor structure or the concurrent validity assessed by correlating a test with other similar tests.

The sensitivity is a psychometric quality that assesses the ability of an evaluation method to detect different enough results among individuals. The sensitivity is assessed by observing the scale scores distribution according to the normal distribution. The sensitivity can be intraindividual or interindividual.

3.3.2 RELIABILITY

We analyzed 116 filled UX in IVE questionnaires. In order to determine the reliability, we calculated the internal consistency of each subscale (i.e. presence, engagement, immersion, flow ...) with Cronbach’s alpha and we calculated the Pearson product-moment correlation coefficient (PCC) of each item through an item analysis. Regarding the Cronbach’s alpha, a value of 0.70 is recommended to consider a measure as being reliable (as a standard/international norm) [7]. We then chose the item analysis to continue with the reliability analysis and check if each item score is consistent compared to the global score of our UX questionnaire. We used Cohen’s convention [5] to interpret the values: a correlation coefficient of 0 indicates that there is no relation. A correlation coefficient of 1 indicates a perfect positive correlation.

A correlation coefficient of -2 indicates a perfect negative correlation. A correlation coefficient between 0.1 and 0.29 indicates a weak correlation. A correlation coefficient between 0.3 and 0.49 indicates a moderate correlation and a correlation coefficient between 0.5 and 1.0 indicates a strong correlation.

3.3.3 SENSITIVITY

In order to determine the sensitivity, we calculated the interindividual sensitivity, that is to say, the scale scores distribution, according to the normal distribution, among different individuals using qualitative observation and the Kolmogorov-Smirnov test (K-S test). The distribution can vary from symmetric distribution (no skew), positively skewed distribution or negatively skewed distribution.

3.3.4 PSYCHOMETRIC PROPERTIES OF OUR UX QUESTIONNAIRE

According to the sample size, we meet the requirements to assess the reliability and the sensitivity of our UX questionnaire. For the validity, we observed that the sample size of our experiment ($N=116$) did not match with the sample size required in the literature ($N=200$ [1], $N = 300$ [22], 10 or more participants per item [19]). Indeed, according to various studies on factor analysis [6] [16] there is an effect of sample size on factor analysis: “*As N increases, sampling error will be reduced and sample factor analysis solutions will be more stable and will more accurately recover the true population structure*”. This does not make possible the validation of our questionnaire with a factor analysis study.

4. RESULTS

We present below the results of the internal consistency with Cronbach’s alpha and an item analysis to check the reliability of our UX questionnaire. In addition, we present the result of the scale scores distribution on distribution graphs and with an analysis of the K-S test to check the interindividual sensitivity of our UX questionnaire. These psychometric properties are calculated with the IBM® SPSS® Statistics software.

4.1 RELIABILITY

4.1.1 INTERNAL CONSISTENCY: SUBSCALE RELIABILITY

Questionnaire reliability data (Cronbach’s alpha α) of our UX questionnaire are presented in *Table 4*. These data indicate for presence, engagement, immersion, flow, skill, emotion, experience consequence, judgement and technology adoption subscales satisfactory internal consistency ($\alpha > 0.7$; i.e. α from 0.718 to 0.908). Whereas for usability subscale the data indicate unsatisfactory internal consistency ($\alpha < 0.7$; i.e. $\alpha = 0.465$).

In analyzing the subscales reliability of our UX questionnaire we found: 11 items (item11 was dropped for a better α for the subscale) for presence provided by Witmer and Singer (1998) with Cronbach's alpha: $\alpha=0.755$ (e.g. *"The virtual environment was responsive to actions that I initiated"*); 3 items for engagement provided by Witmer and Singer (1998) with Cronbach's alpha: $\alpha=0.759$ (e.g. *"The sense of moving around inside the virtual environment was compelling"*); 7 items for immersion provided by Witmer and Singer (1998) with Cronbach's alpha: $\alpha=0.767$ (e.g. *"I felt stimulated by the virtual environment"*); 11 items for flow provided by Heutte (2010) with Cronbach's alpha: $\alpha=0.826$ (e.g. *"I felt I could perfectly control my actions"*); 3 items for usability provided by Brooke (1996) with Cronbach's alpha: $\alpha=0.465$ (e.g. *"I thought the interaction devices -oculus headset, gamepad and/or keyboard- was easy to use"*); 14 items for emotion provided by Pekrun (2011) with Cronbach's alpha: $\alpha=0.718$ (e.g. *"I enjoyed being in this virtual environment"*); 6 items for skill provided by Murphy (1989) with Cronbach's alpha: $\alpha=0.820$ (e.g. *"I felt confident selecting objects in the virtual environment"*); 12 items (grouped in 4) for judgement provided by Hassenzahl, Burmester and Koller (2003) with Cronbach's alpha: $\alpha=0.804$ (e.g. *"Personally, I would say the virtual environment is impractical/practical"*); 9 items for experience consequence provided by Kennedy (1993) with Cronbach's alpha: $\alpha=0.908$ (e.g. *"I suffered from fatigue during my interaction with the virtual environment"*); 9 items for technology adoption provided by Venkatesh (2003) with Cronbach's alpha: $\alpha=0.781$ (e.g. *"If I use again the same virtual environment, my interaction with the environment would be clear and understandable for me"*).

Results (N = 116)		
Subscales	Cronbach's alpha	N° of items
Presence	0.755	11
Engagement	0.759	3
Immersion	0.767	7
Flow	0.826	11
Usability	0.465	3
Skill	0.820	6
Emotion	0.718	14
Experience consequence	0.908	9
Judgement	0.804	12
Technology adoption	0.781	9

Table 4. Results of Cronbach's alpha for our questionnaire's subscales.

4.1.2 INTERNAL CONSISTENCY: ITEM ANALYSIS

Item analysis data of our UX questionnaire are presented in Table 5. The data indicate a satisfactory Pearson product-moment correlation coefficient (PCC) for 68 items out of 87, meaning that these items are significantly correlated (moderately to strongly correlated) with the global score of our UX questionnaire. For 15 items out of 87, the data indicate unsatisfactory PCC, meaning that these items are weakly correlated with the global score of our UX questionnaire.

In analyzing the items correlation of our UX questionnaire we found that: 9 out of 11 items from the presence subscale are significantly correlated with PCC from 0.149 to 0.535; each of the 3 items from the engagement subscale are significantly correlated with PCC from 0.482 to 0.520; 5 out of 7 items from the immersion subscale are significantly correlated with PCC from 0.187 to 0.520; 10 out of 11 items from the flow subscale are significantly correlated with PCC from 0.286 to 0.724; 11 out of 15 items from the emotion subscale are significantly correlated with PCC from 0.055 to 0.737; each of the 6 items from the skill subscale are significantly correlated with PCC from 0.354 to 0.489; 9 out of 12 items from the judgement subscale are significantly correlated with PCC from 0.281 to 0.673;

8 out of 9 items from the experience consequence subscale are significantly correlated with PCC from 0.281 to 0.504; 7 out of 9 items from the technology adoption subscale are significantly correlated with PCC from 0.236 to 0.595.

	<i>PCC</i>		<i>PCC</i>
Item1	0.233*	Item18	0.226*
Item2	0.491**	Item19	0.387**
Item3	0.482**	Item20	0.640**
Item4	0.369**	Item21	0.187*
Item5	0.535***	Item22	0.501***
Item6	0.491**	Item23	0.493**
Item7	0.317**	Item24	0.356**
Item8	0.344**	Item25	0.527***
Item9	0.520***	Item26	0.462***
Item10	0.516***	Item27	0.593***
Item12	0.149*	Item28	0.475**
Item13	0.302**	Item29	0.286*
Item14	0.342**	Item30	0.360**
Item15	0.301**	Item31	0.653***
Item16	0.593***	Item32	0.724***
Item17	0.390***	Item33	0.454**

Table 5. Pearson Correlation Coefficient (PCC) for items 1-38 of our UX questionnaire.

*Weak correlation; ** Moderate correlation; ***Strong correlation

Note. Item 11 was dropped for a better Cronbach's alpha in the presence subscale and items 34, 35, 36 (usability subscale items) were dropped due to unsatisfactory reliability of the usability subscale.

	<i>PCC</i>		<i>PCC</i>
Item37	0.737***	Item64	0.426**
Item38	0.111*	Item65	0.322**
Item39	0.668***	Item66	0.281*
Item40	0.650***	Item52	0.487**
Item41	0.055***	Item53	0.485**
Item42	0.116*	Item54	0.354**
Item43	0.249*	Item55	0.360**
Item44	0.084**	Item56	0.489**
Item45	0.437**	Item57	0.417**
Item46	0.472**	Item58 A	0.673***
Item47	0.458**	Item58 B	0.603***
Item48	0.529***	Item58 C	0.550***
Item49	0.132**	Item67	0.471**
Item50	0.163*	Item68	0.490**
Item51	0.585***	Item69	0.492**
Item59 A	0.281*	Item70	0.445**
Item59 B	0.453**	Item71	0.573***
Item59 C	0.289*	Item72	0.595***
Item60 A	0.367**	Item73	0.567***
Item60 B	0.488**	Item74	0.453**
Item60 C	0.426**	Item75	0.429**
Item61 A	0.461**	Item76	0.459**
Item61 B	0.515***	Item77	0.424**
Item61 C	0.290*	Item78	0.280*
Item62	0.444**	Item79	0.236*
Item63	0.504***		

Table 5. Pearson Correlation Coefficient (PCC) for items 39-79 of our UX questionnaire (Continued).

*Weak correlation; ** Moderate correlation; ***Strong correlation

4.2 SENSITIVITY

Questionnaire sensitivity data of our UX questionnaire are presented in *Table 6* and in *Figures 4-12*. The data indicate for the presence (*Figure 5*), engagement (*Figure 11*), immersion (*Figure 8*), flow (*Figure 4*), emotion (*Figure 10*) and judgement (*Figure 6*) subscales approximatively symmetrically distributed (no skew) scale scores according the normal distribution. Whereas the scale scores of the skill (*Figure 7*), technology adoption (*Figure 9*) and experience consequence (*Figure 12*) subscales are negatively skewed.

The asymmetrical observation of the skill, technology adoption and experience consequence subscales is confirmed by the asymmetry values in *Table 6*. The asymmetry might be explained by the high number of experienced individuals in 3D technologies of our experiment (79 out of 116 participants). Actually, the majority of the participants tend to be skilled in the VE (explained by a majority of positive scores in the skill subscale). They are more likely to adopt 3D technologies as well (explained by a majority of positive scores in the technology adoption subscale) and they tend to be less sick inside the VE (explained by a majority of positive scores in the experience consequence subscale).

Results (N = 116)						
Subscales	Mean	Standard deviations	Variance	Asymmetry	Kolmogorov-Smirnov Z	p
Presence	7.692790	1.0565509	1.116	-0.353	0.839	0.483
Engagement	7.252874	1.6386755	2.685	-0.570	1.062	0.209
Immersion	6.139163	1.7060320	2.911	-0.368	0.740	0.644
Flow	7.349373	1.3249692	1.756	-0.472	0.608	0.853
Skill	8.001437	1.5372187	2.363	-1.390	1.393	0.041
Emotion	6.959770	1.0680744	1.141	-0.062	1.084	0.190
Experience consequence	7.265326	2.3089256	5.331	-0.822	1.809	0.003
Judgement	6.838362	1.4659290	2.149	0.208	0.491	0.969
Technology adoption	7.975096	1.4040878	1.971	-0.925	1.014	0.256

Table 6. Mean and standard deviation of scale scores of our questionnaire.

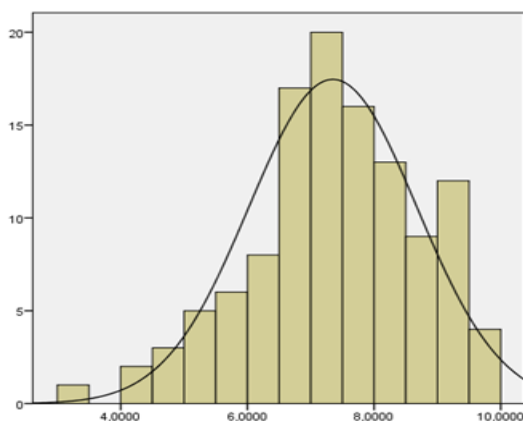


Figure 4. Flow scale scores distribution.

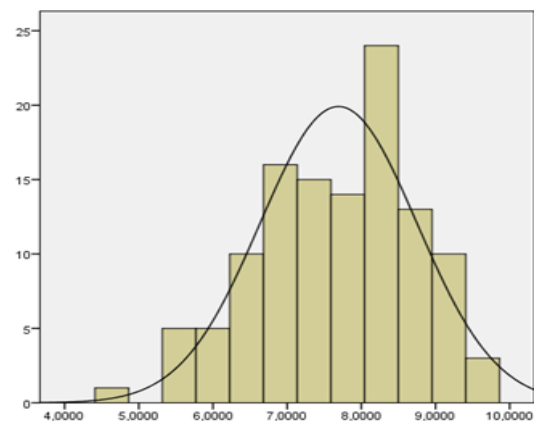


Figure 5. Presence scale scores distribution.

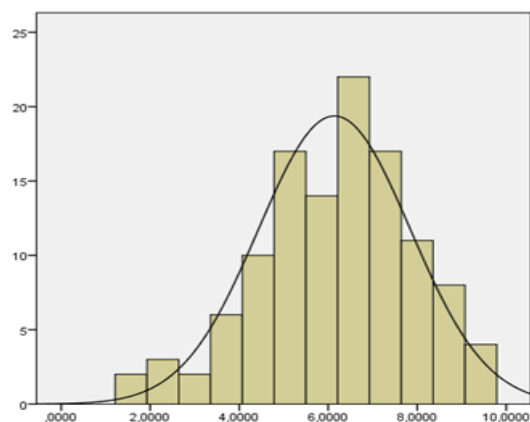


Figure 8. Immersion scale scores distribution.

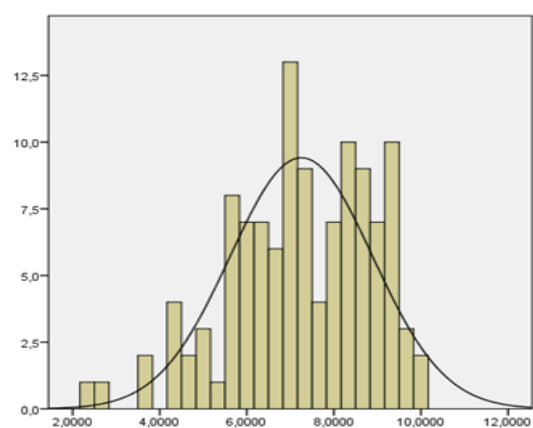


Figure 11. Engagement scale scores distribution.

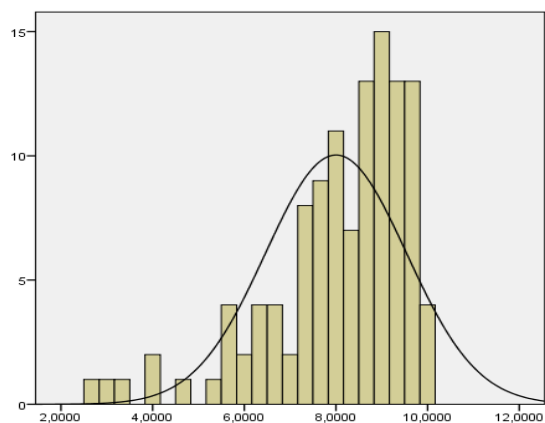


Figure 7. Skill scale scores distribution.

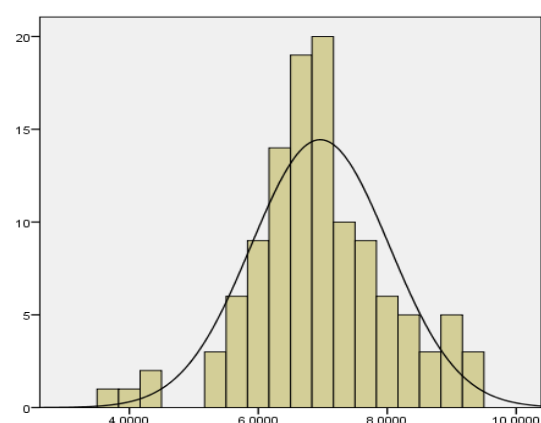


Figure 10. Emotion scale scores distribution.

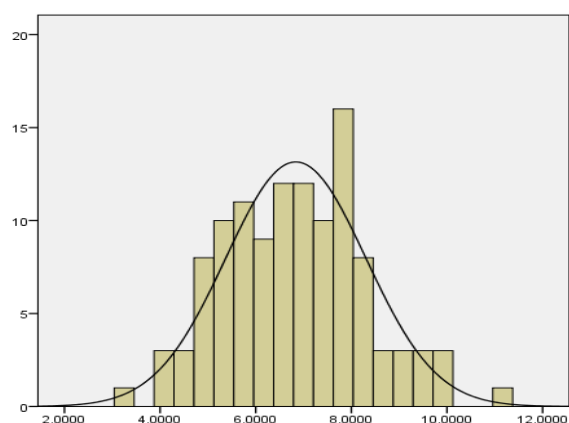


Figure 6. Judgement scale scores distribution.

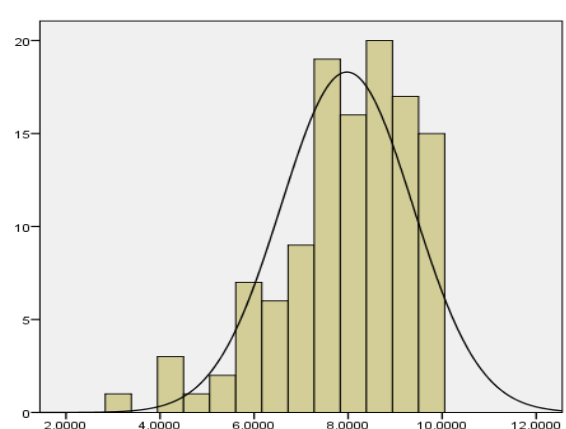


Figure 9. Technology adoption scale scores distribution.

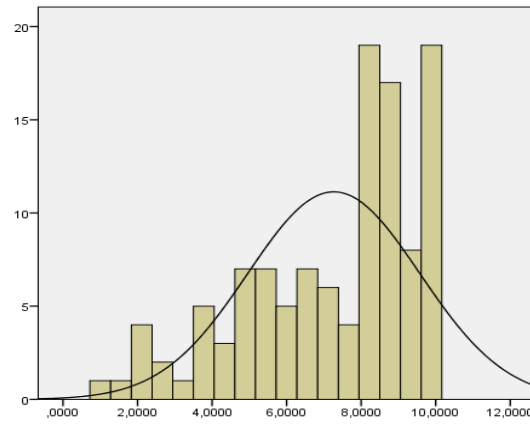


Figure 12. Experience consequence scale scores distribution.

4.3 VALIDATED QUESTIONNAIRE

These data allow us to successfully end with a new version of our validated UX questionnaire according to reliability and sensibility. We dropped 19 items from the original 87 items of our questionnaire. This UX questionnaire is now composed of 68 items and 9 subscales. 9 items compose the presence subscale; 3 items compose the engagement subscale; 5 items compose the immersion subscale; 10 items compose the flow; 11 items compose the emotion subscale; 6 items compose the skill subscale; 9 items compose the judgement subscale; 8 items compose the experience consequence subscale; 7 items compose the technology adoption subscale. See Appendix 1 for the full version of the UX questionnaire.

5. DISCUSSION

This present research set out to integrate the fragmented theory and research on UX into a unified UX in IVE questionnaire. Our study enabled us to assess the reliability of 9 subscales of our UX questionnaire specifically through the internal consistency, provided the rejection of the usability subscale (result show unreliability in the usability subscale with $\alpha = 0.465$ and satisfactory reliability for the other 9 subscales with $\alpha = 0.718 - 0.908$). The usability unsatisfactory result can be explained in several ways: a small amount of items chosen to measure the usability, the items were not enough adjusted to the context or the items chosen were too redundant with other items. In our item analysis, 19 items were dropped because of their low correlation coefficient suggesting that they contributed relatively little to the internal consistency of our questionnaire.

Earlier studies have provided evidence of the internal consistency of the original questionnaires we used to create our own. In comparison in *Table 7*, 7 subscales of our questionnaire have a slightly lower α , 2 subscales have a slightly higher α and one subscale have a strongly lower α . Our findings concerning the reliability of our questionnaire are in certain aspects similar to the original studies (*Table 7*). This might suggest that the selection criteria of the questionnaires and items contributed to these encouraging results.

Component	Original Questionnaire	α in our study	α in original questionnaire	Authors of the questionnaire study
Presence	PQ	0.755	0.88	Kennedy et al. 1993 [13]
Engagement	PQ	0.759		
Immersion	ITQ	0.767	0.81	Heutte et al. 2010 [12]
Flow	Flow4D16	0.826	0.84 - 0.86	
Usability	SUS	0.465	0.92	Lewis et al. 2009 [15]
Skill	CSE	0.820	0.95	Murphy et al. 1989 [18]
Emotion	AEQ	0.718	0.78 - 0.93	Pekrun et al. 2011 [21]
Experience consequence	SSQ	0.908	0.71	Bailenson et al. 2006 [2]
Judgement	AttracDiff	0.804	0.73 - 0.90	Hassenzahl et al. 2003 [11]
Technology adoption	UTAUT	0.781	0.87 - 0.91	Venkatesh et al. 2003 [25]

Table 7. Comparison of alphas in the original questionnaires and in our UX in IVE questionnaire.

So far, only a limited number of studies investigated the sensitivity of a UX questionnaire. We assessed the sensitivity through observation and the K-S test and found scale scores normal distribution for 6 subscales (i.e. presence, engagement, immersion, flow, emotion, judgement), and negative skewed distribution for 3 subscales (i.e. skill, technology adoption and experience consequence). These 3 subscales negatively skewed can be explained by a high number of experienced individuals in 3D technologies that tend to be skilled in the VE, that are likely to adopt 3D technologies more easily and that tend to be less sick inside the VE.

This UX questionnaire is a non-definitive tool and still needs adjustments, it is then necessary to recognize some of our UX questionnaire's limitations. One limitation concerns the lack of investigation of other reliability parameters such as the test-retest method, due to the already big workload for the participants requested by our experimental protocol. Indeed, the load requested by the experiment might have caused more risks of errors and fatigue in the experiment process and more risk of random answers in the questionnaire. Nevertheless, measuring the reliability along time is feasible now that our questionnaire does not need so much validation. We might consider an experiment with participants during two days, the exact same experiment with the exact same participants will take place on the first and the second day. Another limitation concerns the lack of investigation of other sensitivity parameters such as the intraindividual sensitivity for the same reason as the test-retest method. A further limitation concerns the lack of investigation of the validity parameters such as the construct validity due to the unsatisfactory sample size required or the criterion validity due to the incompatibility with our experimental protocol. Indeed, the criterion validity method requests a comparison between our unified questionnaire and the original questionnaires. Unfortunately, no dataset of the original questionnaires was collected because the experiment would have been very cumbersome for the participants (i.e. the experiment would have requested a completion of 10 questionnaires for each participant).

6. CONCLUSION

The work discussed in this paper provides a method to validate a UX questionnaire for IVEs. More specifically, we validate a UX questionnaire subscales for an edutainment IVE. This work provides a measurement tool, aiming to measure the multiple facets of UX in IVE in the edutainment field. We were able to prove the quality of our questionnaire: for the reliability, we were able to validate the internal consistency with Cronbach's alpha (except for one subscale) and item correlation that reinforces the reliability of our UX questionnaire subscales (provided 19 items dropped). For the sensitivity, we observed the normal distribution of the scale scores for 6 subscales and the negative skewed distribution of the scale scores for 3 subscales in our UX questionnaire, due to the high number of skilled participants with VE.

Some questions remain on the unsatisfactory internal consistency result of the usability subscale, and on the validity analysis of our UX questionnaire. Obviously, we decided to follow with the next steps of the validation process without taking into account the usability subscale, that is to say the subjective usability, as it is not reliable. We consider our validation process suitable for experiments with less than 200 participants, but we are aware that a higher number of participants would have helped enrich and refine the whole validation process (i.e. construct validity, concurrent validity, confirmatory factor analysis).

This study provides important new insight into UX in IVE assessment. The validation process of our questionnaire may be extended to different types of VE (in this study we used an edutainment VE) such as therapeutic, design or collaborative applications. Moreover, the process can be extended to different types of devices (in this study we used a HMD) such as a Cave Automatic Virtual Environment (CAVE), Z-space, ... Given our holistic UX in IVE questionnaire, we expect our questionnaire to be validated in several fields, with different applications and Virtual Reality (VR) technologies.

Finally, the present study has a number of important implications for UX design. First, our UX in IVE questionnaire can be used in the earlier phases of a VE design. The questionnaire might be used as soon as a prototype is designed to assess the UX and therefore favor a better UX for the final product. Secondly, designers can use the UX in IVE questionnaire as a guide, to help them focus on the unsatisfactory aspects of the VE in terms of UX, knowing that way, which aspect of the UX has to be improved to provide a greater and suitable experience for customers.

7. ACKNOWLEDGEMENTS

Our thanks go to Laval Agglomération for their financial support and to the participants who took time to engage and provide feedback in this project.

8. REFERENCES

- [1] Anderson, J. C., & Gerbing, D. W., 1984. The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49(2), 155-173.
- [2] Bailenson, J. N., & Yee, N., 2006. A longitudinal study of task performance, head movements, subjective report, simulator sickness, and transformed social interaction in collaborative virtual environments. *Presence: Teleoperators and Virtual Environments*, 15(6), 699-716.
- [3] Brooke, J., 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4-7

- [4] Chai, J., Ge, Y., Liu, Y., Li, W., Zhou, L., Yao, L., Sun, X., 2014. Application of frontal EEG asymmetry to user experience research. In: Harris, D. (ed.) EPCE 2014. LNCS, vol. 8532, pp. 234–243. Springer, Heidelberg
- [5] Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences. 2nd edn. Hillsdale, New Jersey: L.
- [6] Comrey, A. L., & Lee, H. B., 1992. A first course in factor analysis. Hillsdale, NJ: Erlbaum
- [7] Devellis, R. F., 2003. Scale Development: Theory and Applications, Second Edition, Applied Social Research Methods Series, 26, Thousand Oaks, Calif.: Sage
- [8] Fernandez L., Aulagnier M., Bonnet A., Guinard A., Pedinielli J.L., Préau M., 2005. Module VI : Outils Psychométriques, dans P. Verger, M. Aulagnier, V. Schwœbel & T. Lang (Dir), Démarches épidémiologiques après une catastrophe. Online : [http://www.invs.sante.fr/publications/2005/epidemiologie catastrophes/module6.pdf](http://www.invs.sante.fr/publications/2005/epidemiologie%20catastrophes/module6.pdf)
- [9] Georges, V., Courtemanche, F., Senecal, S., Baccino, T., Fredette, M., & Leger, P. M., 2016. UX Heatmaps: Mapping User Experience on Visual Interfaces. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 4850-4860). ACM.
- [10] Hassan, S. E., Hicks, J. C., Lei, H., & Turano, K. A., 2007. What is the minimum field of view required for efficient navigation? Vision research, 47(16), 2115-2123.
- [11] Hassenzahl, M., Burmester, M., & Koller, F., 2003. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität [AttracDiff: A questionnaire to measure perceived hedonic and pragmatic quality]. In J. Ziegler & G. Szwillus (Eds.), Mensch & Computer 2003. Interaktion in Bewegung (pp. 187–196). Stuttgart, Leipzig: B. G. Teubner.
- [12] Heutte, J., & Fenouillet, F., 2010. Propositions pour une mesure de l'expérience optimale (état de Flow) en contexte éducatif. Actes du 26e congrès international d'actualité de la recherche en éducation et en formation (AREF).
- [13] Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G., 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. The international journal of aviation psychology, 3(3), 203-220.
- [14] Kim, C. J., & Chang, M. H., 2015. Actual Emotion and False Emotion Classification by Physiological Signal. In 2015 8th International Conference on Signal Processing, Image Processing and Pattern Recognition (SIP) (pp. 21-24). IEEE.
- [15] Lewis, J. R., & Sauro, J., 2009. The factor structure of the system usability scale. In Human centered design (pp. 94-103). Springer Berlin Heidelberg.
- [16] MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S., 1999. Sample size in factor analysis. Psychological methods, 4(1), 84.
- [17] McFarland, D. J., Sarnacki, W. A., Vaughan, T. M., & Wolpaw, J. R., 2005. Brain-computer interface (BCI) operation: signal and noise during early training sessions. Clinical Neurophysiology, 116, 56-62.
- [18] Murphy, C. A., Coover, D., & Owen, S. V., 1989. Development and validation of the computer self-efficacy scale. Educational and Psychological measurement, 49(4), 893-899.
- [19] Nunnally, J. C., 1978. Psychometric Theory, (2nd ed.) New York: McGraw-Hill.
- [20] Pallot, M., Eynard, R., Poussard, B., Christmann, O. and Richir, S., 2013. Augmented sport: exploring collective user experience. In Proceedings of the Virtual Reality International Conference: Laval Virtual (VRIC '13). ACM, New York, NY, USA, Article 4, 8 pages. DOI=10.1145/2466816.2466821 <http://doi.acm.org/10.1145/2466816.2466821>

- [21] Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P., 2011. Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, 36(1), 36-48.
- [22] Tabachnik, B. G., & Fidell, S. L., 2007. *Discriminant analysis. Using multivariate statistics*. Boston: Pearson Education Inc, 201(3), 377-438.
- [23] Tcha-Tokey, K., Loup-Escande, E., Christmann, O., Canac, G., Farin, F., & Richir, S., 2015. Towards a user experience in immersive virtual environment model: a review. In *Proceedings of the 27th Conference on l'Interaction Homme-Machine* (p. 26). ACM.
- [24] Tcha-Tokey, K., Loup-Escande, E., Christmann, O., & Richir, S., 2016. A questionnaire to measure the User Experience in Immersive Virtual Environment. In *Proceedings of the 18th VRIC 2016*. ACM.
- [25] Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D., 2003. User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.
- [26] Witmer, B. G., & Singer, M. J., 1998. Measuring presence in virtual environments: A presence questionnaire, *Presence* 7(3): 225-240
- [27] Yao, L., Liu, Y., Li, W., Zhou, L., Ge, Y., Chai, J., & Sun, X., 2014. Using physiological measures to evaluate user experience of mobile applications. In *Engineering Psychology and Cognitive Ergonomics* (pp. 301-310). Springer International Publishing.
- [28] Oculus Best Practices (n.d) Retrieved from https://developer.oculus.com/documentation/intro-vr/latest/concepts/bp_intro/

APPENDIX 1.

English translation of our unified UX in IVE questionnaire (originally in French).

Items	Subscale
1. The virtual environment was responsive to actions that I initiated.	<i>Presence</i>
2. My interactions with the virtual environment seemed natural.	<i>Presence</i>
3. The visual aspects of the virtual environment involved me.	<i>Engagement</i>
4. The devices (gamepad or keyboard) which controlled my movement in the virtual environment seemed natural.	<i>Presence</i>
5. I was able to actively survey the virtual environment using vision.	<i>Presence</i>
6. The sense of moving around inside the virtual environment was compelling.	<i>Engagement</i>
7. I was able to examine objects closely.	<i>Presence</i>
8. I could examine objects from multiple viewpoints.	<i>Presence</i>
9. I was involved in the virtual environment experience.	<i>Engagement</i>
10. I felt proficient in moving and interacting with the virtual environment at the end of the experience.	<i>Presence</i>
11. The visual display quality distracted me from performing assigned tasks.	<i>Presence</i>
12. The devices (gamepad or keyboard) which controlled my movement distract me from performing assigned tasks.	<i>Presence</i>
13. I could concentrate on the assigned tasks rather than on the devices (gamepad or keyboard).	<i>Presence</i>
14. I correctly identified sounds produced by the virtual environment.	<i>Presence</i>
15. I correctly localized sounds produced by the virtual environment.	<i>Presence</i>
16. I felt stimulated by the virtual environment.	<i>Immersion</i>
17. I become so involved in the virtual environment that I was not aware of things happening around me.	<i>Immersion</i>
18. I identified to the character I played in the virtual environment.	<i>Immersion</i>
19. I become so involved in the virtual environment that it is if I was inside the game rather than manipulating a gamepad and watching a screen.	<i>Immersion</i>
20. I felt physically fit in the virtual environment.	<i>Immersion</i>
21. I got scared by something happening in the virtual environment.	<i>Immersion</i>
22. I become so involved in the virtual environment that I lose all track of time.	<i>Immersion</i>
23. I felt I could perfectly control my actions.	<i>Flow</i>
24. At each step, I knew what to do.	<i>Flow</i>
25. I felt I controlled the situation.	<i>Flow</i>
26. Time seemed to flow differently than usual.	<i>Flow</i>

Items (Continued)	Subscale
27. Time seemed to speed up.	<i>Flow</i>
28. I was losing the sense of time.	<i>Flow</i>
29. I was not worried about other people's judgement.	<i>Flow</i>
30. I was not worried about what other people would think of me.	<i>Flow</i>
31. I felt I was experiencing an exciting moment.	<i>Flow</i>
32. This experience was giving me a great sense of well-being	<i>Flow</i>
33. When I mention the experience in the virtual environment, I feel emotions I would like to share.	<i>Flow</i>
34. I thought the interaction devices (Oculus headset, gamepad and/or keyboard) was easy to use.	<i>Usability</i>
35. I thought there was too much inconsistency in the virtual environment.	<i>Usability</i>
36. I found the interaction devices (Oculus headset, gamepad and/or keyboard) very cumbersome to use.	<i>Usability</i>
37. I enjoyed being in this virtual environment.	<i>Emotion</i>
38. I got tense in the virtual environment.	<i>Emotion</i>
39. It was so exciting that I could stay in the virtual environment for hours.	<i>Emotion</i>
40. I enjoyed the experience so much that I feel energized.	<i>Emotion</i>
41. I felt nervous in the virtual environment.	<i>Emotion</i>
42. I got scared that I might do something wrong.	<i>Emotion</i>
43. I worried whether I was able to cope with all the instructions that was given to me.	<i>Emotion</i>
44. I felt like distracting myself in order to reduce my anxiety.	<i>Emotion</i>
45. I found my mind wandering while I was in the virtual environment.	<i>Emotion</i>
46. The interaction devices (Oculus headset, gamepad and/or keyboard) bored me to death.	<i>Emotion</i>
47. When my actions were going well, it gave me a rush.	<i>Emotion</i>
48. While using the interaction devices (Oculus headset, gamepad and/or keyboard), I felt like time was dragging.	<i>Emotion</i>
49. I enjoyed the challenge of learning the virtual reality interaction devices (Oculus headset, gamepad and/or keyboard)	<i>Emotion</i>
50. The virtual environment scared me since I do not fully understand it.	<i>Emotion</i>
51. I enjoyed dealing with the interaction devices (Oculus headset, gamepad and/or keyboard).	<i>Emotion</i>
52. I felt confident selecting objects in the virtual environment.	<i>Skill</i>
53. I felt confident moving the cross hair around the virtual environment.	<i>Skill</i>
54. I felt confident using the gamepad and/or keyboard to move around the virtual environment.	<i>Skill</i>

Items (Continued)	Subscale
55. I feel confident understanding the terms/words relating to the interaction devices (Oculus headset, gamepad and/or keyboard).	<i>Skill</i>
56. I feel confident learning advanced skills within a specific virtual reality software using the Oculus headset.	<i>Skill</i>
57. I feel confident describing the functions the interaction devices (Oculus headset, gamepad and/or keyboard, joystick, Kinect ...) of a virtual reality environment.	<i>Skill</i>
58. A. Personally, I would say the virtual environment is impractical/practical	<i>Judgement</i>
58. B. Personally, I would say the virtual environment is confusing/clear	<i>Judgement</i>
58. C. Personally, I would say the virtual environment is unruly/manageable	<i>Judgement</i>
59. A. I found that this virtual environment was typical/original	<i>Judgement</i>
59. B. I found that this virtual environment was lame/exciting	<i>Judgement</i>
59. C. I found that this virtual environment was easy/challenging	<i>Judgement</i>
60. A. I found this virtual environment amateurish/professional	<i>Judgement</i>
60. B. I found this virtual environment gaudy/classy	<i>Judgement</i>
60. C. I found this virtual environment unpresentable/presentable	<i>Judgement</i>
61. A. I found that this virtual environment is ugly/beautiful	<i>Judgement</i>
61. B. I found that this virtual environment is disagreeable/likeable	<i>Judgement</i>
61. C. I found that this virtual environment is discouraging/motivating	<i>Judgement</i>
62. I suffered from fatigue during my interaction with the virtual environment.	<i>Experience consequence</i>
63. I suffered from headache during my interaction with the virtual environment.	<i>Experience consequence</i>
64. I suffered from eyestrain during my interaction with the virtual environment.	<i>Experience consequence</i>
65. I felt an increase of my salivation during my interaction with the virtual environment.	<i>Experience consequence</i>
66. I felt an increase of my sweat during my interaction with the virtual environment.	<i>Experience consequence</i>
67. I suffered from nausea during my interaction with the virtual environment.	<i>Experience consequence</i>
68. I suffered from “fullness of the head” during my interaction with the virtual environment.	<i>Experience consequence</i>
69. I suffered from dizziness with eye open during my interaction with the virtual environment.	<i>Experience consequence</i>
70. I suffered from vertigo during my interaction with the virtual environment.	<i>Experience consequence</i>

Items (Continued)	Subscale
71. If I use again the same virtual environment, my interaction with the environment would be clear and understandable for me.	<i>Technology adoption</i>
72. It would be easy for me to become skillful at using the virtual environment.	<i>Technology adoption</i>
73. Learning to operate the virtual environment would be easy for me	<i>Technology adoption</i>
74. Using the interaction devices (Oculus headset, gamepad and/or keyboard) is a bad idea.	<i>Technology adoption</i>
75. The interaction devices (Oculus headset, gamepad and/or keyboard) would make work more interesting.	<i>Technology adoption</i>
76. I would like working with the interaction devices (Oculus headset, gamepad and/or keyboard).	<i>Technology adoption</i>
77. I have the resources necessary to use the interaction devices (Oculus headset, gamepad and/or keyboard).	<i>Technology adoption</i>
78. I have the knowledge necessary to use the interaction devices (Oculus headset, gamepad and/or keyboard).	<i>Technology adoption</i>
79. The interaction devices (Oculus headset, gamepad and/or keyboard) are not compatible with other technologies I use.	<i>Technology adoption</i>
80. In your opinion, what were the positive points about your experience?	<i>None</i>
81. In your opinion, what were the negative points about your experience?	<i>None</i>
82. Do you have suggestions to improve this virtual reality environment?	<i>None</i>